

**Amendments to the Specification:**

Please replace paragraphs [0006], [0038]-[0048], [0050], [0053]-[0059], and [0063] with the following amended paragraphs:

[0006] The simplest method of transcoding is a brute-force approach called tandem transcoding, shown in Figure 1. This method performs a full decode 110 of the incoming compressed bits to produce synthesized speech 112. The synthesized speech is then encoded 114 for the target standard. This method is undesirable because of the huge amount of computation performed in re-encoding the signal, as well as quality degradations introduced by pre- and post-filtering of the speech waveform, and the potential delays introduced by the look-ahead-requirements of the encoder.

[0038] A block diagram of a tandem connection between two voice codecs 110, 114 is shown in Figure 1. Alternatively a transcoder 210 may be used, as shown in Figure 2, which converts the bitstream from a source codec to the bitstream of a destination codec without fully decoding the signal to PCM and then re-encoding the signal. The present invention is a transcoder between voice codecs, whereby the destination codec is a variable bit-rate voice codec that determines the bit-rate based on the input speech characteristics. A block diagram of the encoder of a variable bit-rate voice coder is shown in Figure 3. The input speech signal passes through several processing stages including pre-processing 310, estimation of model parameters 320 and computation of classification features 322. Then, a rate, and in some cases, a frame type, is determined based on the features detected 324. Depending on the rate decision, a different strategy may be used in the encoding process 330, 332. Once coding is complete, the parameters are packed in the bitstream 340.

[0039] A diagram of the apparatus for transcoding between two variable bit-rate ~~voice~~ voice codecs of the present invention is shown in Figure 4. The apparatus comprises a source codec unpacking module 410, an intermediate parameters interpolation module 420, a smart frame classification and rate determination module 422, several mapping strategy modules 430, 432, a switching module 450 to select the desired mapping strategy, a destination packet formation

module 440, and a second switching module 452 that links the mapping strategy to the destination packet formation module 440. The method for transcoding between two variable bit-rate ~~voice~~ voice codecs is shown in Figure 5.

[0040] Firstly, the bitstream representing frames of data encoded according to the source voice codec is unpacked and unquantized by a bitstream unpacking module 410. The actual parameters extracted from the bitstream depend on the source codec and its bit rate, and may include line spectral frequencies, pitch delays, delta pitch delays, adaptive codebook gains, fixed codebook shapes, fixed codebook gains and frame energy. Particular voice codecs may also transmit information regarding spectral transition, interpolation factors, the switch predictor used as well as other minor parameters. The unquantised parameters are passed to the intermediate parameters interpolation module 420.

[0041] The intermediate parameters interpolation module 420 interpolates between different frame sizes, subframe sizes and sampling rates. This is required if there are differences in the frame size or subframe size of the source and destination codecs, in which case the transmission frequency of parameters may not be matched. Also, a difference in the sampling rate between the source codec and destination codec requires modification of parameters. The output interpolated parameters 402 are passed to the smart frame classification and rate determination module and one of the mapping modules 422.

[0042] The frame classification and rate determination module 422 receives the unquantized interpolated parameters of the source codec 402 and the external control commands of the destination codec 404, as shown in Figure 6. The frame classification and rate determination module 422 comprises a classifier input parameter selector, for selecting which inputs will be used in the classification task, M sub-classifiers, buffers to store past input parameters and past output values, and a final decision module. The classifier takes as input the selected classification input parameters 402, external commands 404, and past input and output values 602, and generates as output the frame class and rate decision 406 for the destination codec. Once classification has been performed, the states of the data buffers storing past parameter values are updated 610. The output rate and frame type decision 406 controls the first switching module 450 that selects the parameter mapping module, and the second switching module 452

that links the parameter mapping module to the bitstream packing module 440. ~~frame~~ Frame classification is performed according to pre-defined coefficients or rules determined during a prior training or classifier construction process. Several types of classification techniques may be used, including but not exclusive to, decision trees, rule-based models, and artificial neural networks. The functions for computing classification features and the many steps of the classification procedure for a particular codec are shown in Figure 7 and Figure 8 respectively. In an embodiment of the present invention, the frame classification and rate determination module replaces the standard classifier of the destination codec, as well as the processing functions of the destination codec required to generate the classification parameters.

[0043] The intermediate parameters interpolation module 420 and the frame classification and rate determination module 422 are linked to one of many parameter mapping modules 430, 432 by a switching module 450. The destination codec frame type and bit rate determined 406 by the frame classification and rate determination module 422 control which mapping module is to be chosen 422. Mapping modules 430, 432 may exist for each combination of bit-rate and frame class of the source codec to each bit rate and frame class of the destination codec.

[0044] Each mapping module comprises a speech spectral parameter mapping unit 910, an excitation mapping unit 920, and a mapping strategy decision unit 930. The speech spectral parameter mapping unit 910 maps the spectral parameters, usually line spectral pairs (LSPs) or line spectral frequencies (LSFs), of the source codec 911, directly to the spectral parameters of the destination codec 912. A calibration factor 914 is calculated and used to calibrate the excitation to account for the differences in the quantised spectral parameters of the source and destination codec. The excitation mapping unit 920 takes CELP excitation parameters including pitch lag, adaptive codebook gain, fixed codebook gain and fixed codebook codevectors from the interpolator and maps these to encoded CELP excitation parameters according to the destination codec. Figure 9 shows a mapping module which may be selected for mapping parameters of an active speech frame, e.g., mapping from Rate  $\frac{1}{2}$  or Rate 1 of EVRC to Rate  $\frac{1}{2}$  or Rate 1 of SMV. In this case, the input parameters to the excitation coding mapping unit are the adaptive codebook lag 921, adaptive codebook gain 923, fixed codebook codevector 927 and fixed codebook gain 925 of the source codec. The output parameters to the excitation coding mapping

unit are the adaptive codebook lag 922, adaptive codebook gain 924, fixed codebook codevector 928 and fixed codebook gain 926 in the format of the destination codec. Figure 10 shows a mapping module 1000 which may be selected for mapping parameters of a silence or noise-like speech frame, e.g., mapping from Rate 1/8 of EVRC to Rate 1/4 or Rate 1/8 of SMV. In this case, the input parameters to the excitation coding mapping unit 1020 are typically the frame energy or subframe energies 1021, and excitation shape 1023. Not all excitation parameters shown in the figures may be present for a given codec or bit rate.

[0045] Linked to the excitation coding mapping unit 920 is a mapping strategy decision unit 930, which controls the type of excitation mapping to be used. Several mapping approaches may be used, including those using direct mapping from source codec to destination codec without any further analysis or iterations, analysis in the excitation domain, analysis in the filtered excitation domain or a combination of these strategies, such as searching the adaptive codebook in the excitation space and fixed codebook in the filtered excitation space. The mapping strategy decision module determines which mapping strategy is to be applied. The decision may be based on available computational resources or minimum quality requirements and can change in a dynamic fashion.

[0046] Except for the direct mapping strategy, in which parameters are directly mapped from source codec format to destination codec format without any analysis, the excitation signal is reconstructed. Reconstruction of the excitation during active speech requires the interpolated excitation parameters of pitch delays, adaptive codebook gains, fixed codebook shapes, and fixed codebook gains. During silence or noise, the parameters required are the signal energy, signal shape if available, and a random noise generator. Figure 11 shows a block diagram the decoding process performed in a RCELP-based voice decoder. In this figure, the linear prediction (LP) excitation is formed by combining the gain-scaled contributions of the adaptive and fixed codebooks 1120, 1122, and then filtered by the speech synthesis filter 1124 and post-filter 1130. In the transcoder architecture of the present invention, to reduce complexity and quality degradations, the final source codec decoder operations of filtering the LP excitation signal by the synthesis filter to convert to the speech domain and then post-filtering to mask quantization noise are not used. Similarly, the pre-processing operations in the encoder of the destination

codec are not used. An example of a speech pre-processor is shown in Figure 12. High-pass filtering 1212 is a common pre-processing step in existing CELP-based voice codecs, with the advanced steps of silence enhancement 1210, noise suppression 1214 and adaptive tilt filtering 1216 being applied in more recent voice codecs. In the case where the source codec does not use noise suppression and the destination codec does use noise suppression, the transcoder architecture should provide noise suppression functionality.

[0047] Current variable-rate voice codecs applicable to the present invention include EVRC and SMV which are based on the Relaxed CELP (RCELP) principle. Typical excitation quantization in RCELP codecs is performed by the technique shown in Figure 13 and Figure 14. In this case, the target signal is modified weighted speech 1302. The modification is performed to create a signal with a smooth interpolated pitch delay contour by time-warping or time-shifting pitch pulses. This allows for coarse pitch quantization. The adaptive codebook 1310 is mapped to the delay contour and then searched by gain-adjusting 1320 and filtering each candidate vector by the weighted synthesis filter 1330, 1340 and comparing the result to the target signal 1302. Once the best adaptive codebook vector is found, its contribution is subtracted from the target 1350, and the fixed codebook 1360 is searched in a similar manner. In the case where both source and destination codecs are based on the RCELP principle, the computationally expensive operation of detecting and shifting each pitch pulse in the encoder processing of the destination codec is not required. This is due to the fact that the reconstructed source excitation already follows the interpolated pitch track of the source codec. Hence, the target signal in the transcoder is not modified weighted speech, but simply the weighted speech, speech, weighted excitation, excitation, or calibrated excitation signal.

[0048] Figure 15 shows a block diagram of an example of one mapping strategy of the transcoder between variable-rate voice codecs of the present invention. The procedure is outlined in Figure 16. In this case, the mapping strategy chosen is a combination between analysis in the excitation domain and analysis in the filtered excitation domain. The target signal for the adaptive codebook search is the calibrated excitation signal 1502. The search of the adaptive codebook 1510 is performed in the excitation domain. This reduces complexity as each candidate codevector does not need to be filtered with the weighted synthesis filter before it can

be compared to a speech domain target signal. The initial estimate of the pitch lag is the pitch lag obtained from the interpolation module that has been interpolated to match the subframe size of the destination codec 1610. The pitch is searched within a small interval of the initial pitch estimate 1612, at the accuracy (integer or fractional pitch) required by the destination codec. The adaptive codebook gain is then determined for the best codevector 1614 and the adaptive codevector contribution is removed from the calibrated excitation 1616. The result is filtered using a special weighting filter to produce the target signal for the fixed codebook search 1618. The fixed codebook is then searched, either by a fast technique or by gain-adjusting and filtering candidate codevectors by the special weighting filter and comparing the result with the target 1620, 1622, 1624. Fast search methods may be applied for both the adaptive and fixed codebook searches.

[0050] A second-stage switching module 452 links the interpolation and mapping module to the destination bitstream packing module 440. The destination bitstream packing module 440 packs the destination CELP parameters in accordance with the destination codec standard. The parameters to be packed depend on the destination codec, the bit rate and frame type.

#### EVRC ⇔ SMV TRANSCODING EXAMPLE

[0053] A diagram of the apparatus for transcoding from EVRC to SMV is shown in Figure 17. The apparatus comprises an EVRC unpacking module 1710, an intermediate parameters interpolation module 1720, a smart SMV frame classification and rate determination module 1730, several mapping modules 1740, 1742, 1744, 1746 to map parameters from all allowed rate and type transcoder transitions, and a SMV packet formation module 1750. The inputs to the apparatus are the EVRC frame packets 1702 and SMV external commands 1704 (e.g. network-controlled mode, half-rate max flag), and the outputs are the SMV frame packets 1706. Similarly, the apparatus for transcoding from SMV to EVRC is shown in Figure 18. The apparatus comprises a SMV unpacking module 1810, an intermediate parameters interpolation module 1820, an EVRC rate determination module 1830, several mapping modules 1840, 1842, 1844, 1846 to map parameters from all allowed rate and type transcoder transitions, and an EVRC packet formation module 1850. The inputs to the apparatus are the SMV frame packets

1802 and EVRC external commands 1804 (e.g. half-rate max flag), and the outputs are the EVRC frame packets 1806.

[0054] In transcoding from EVRC to SMV, the bitstream representing frames of data encoded according to EVRC is unpacked by a bitstream unpacking module 1710. The actual parameters from the bitstream depend on the EVRC bit rate and include line spectral frequencies, spectral transition indicator, pitch delay, delta pitch delay, adaptive codebook gain, fixed codebook shapes, fixed codebook gains and frame energy. The unquantised parameters are passed to the intermediate parameters interpolation module 1720.

[0055] The intermediate parameter interpolation module 1720 interpolates between the different subframe sizes of EVRC and SMV. EVRC has 3 subframes per frame, whereas SMV has 1, 2, 3, 4, or 10 subframes per frame depending on the bit rate and frame type. Depending on the parameter and coding strategy, subframe interpolation may or may not be required. Figure 19 and Figure 20 illustrate the frame and subframe sizes for the different rates and frame types of SMV and EVRC respectively. Since the frame size of both codecs is 20ms and the sampling rate of both codecs is 8kHz, no frame size or sampling rate interpolation is required. The output interpolated parameters, or if no interpolation was carried out, the EVRC CELP parameters, are passed to the smart frame classification and rate determination module and the selected of the mapping module.

[0056] The frame classification and rate determination module 1730 receives the EVRC CELP parameters 1712, the EVRC bit rate 1714, the SMV network-controlled mode and any other SMV external commands 1704. The frame classification and rate determination module 1730 produces a frame class and rate decision 1716 for SMV based on these inputs. The frame classification and rate determination module 1730 comprises a classifier input parameter selector, for selecting which of the EVRC parameters will be used as inputs to the classification task, M sub-classifiers, buffers to store past input parameters and past output values and a final decision module. The sub-classifiers take as input the selected classification input parameters, the SMV network-controlled mode command, and past input and output values, and generate the frame class and rate decision. One sub-classifier may be used to determine the bit rate, and a second sub-classifier may be used to determine the frame class. The SMV frame class is either

silence, noise-like, unvoiced, onset, non-stationary voiced or stationary voiced, and the SMV rate may be Rate 1, Rate  $\frac{1}{2}$ , Rate  $\frac{1}{4}$ , or Rate  $\frac{1}{8}$ . The SMV frame classification, using EVRC parameters, is performed according to a pre-defined configuration and classifier algorithm. The coefficients or rules of the classifier are determined during a prior EVRC-to-SMV classifier training or construction process. The frame classification and rate determination module includes a final decision module, that enforces all SMV rate transition rules to ensure illegal rate transitions are not allowed. For example, in SMV, a Rate 1 Type 1 cannot follow a Rate  $\frac{1}{8}$  frame. This frame classification and rate determination module replaces the SMV standard classifier, which requires a large amount of processing to derive the parameters and features required for classification. The SMV frame-processing functions are shown in Figure 7, and the many steps of the SMV classification procedure are shown in Figure 8. These functions are not necessary in the present invention as the already available EVRC CELP parameters are used as inputs to classifier module.

[0057] The intermediate parameters interpolation module 1720 and the SMV smart frame classification and rate determination module 1730 are linked to one of many interpolation and mapping modules 1740, 1742, 1744, 1746 by a switching module 1760. EVRC has a single processing algorithm for each rate, whereas SMV has two possible processing algorithms for each of Rate 1 and Rate  $\frac{1}{2}$ , and a single processing algorithm for each of Rate  $\frac{1}{4}$  and Rate  $\frac{1}{8}$ . The SMV frame type and bit rate 1716 determined by the frame classification and rate determination module control which interpolation and mapping module is to be chosen. For Rates 1 and  $\frac{1}{2}$  of SMV, the stationary voiced frame class uses subframe processing Type 1 and all other frame classes use subframe processing Type 0. As shown in Figure 17, there are interpolation and mapping modules 1740, 1742, 1744, 1746 for each allowed EVRC rate and SMV type and rate combination. For example, interpolation and mapping modules include:

EVRC Rate 1 to SMV Rate 1 Type 0

EVRC Rate 1 to SMV Rate 1 Type 1

EVRC Rate  $\frac{1}{2}$  to SMV Rate 1 Type 0

EVRC Rate  $\frac{1}{2}$  to SMV Rate 1 Type 1



EVRC Rate  $\frac{1}{2}$  to SMV Rate  $\frac{1}{2}$  Type 0

EVRC Rate  $\frac{1}{2}$  to SMV Rate  $\frac{1}{2}$  Type 1

....

and so on.

[0058] For the EVRC-to-SMV transcoder, interpolation and mapping modules 1840, 1842, 1844, 1846 include:

SMV Rate 1 Type 0 to EVRC Rate 1

SMV Rate 1 Type 1 to EVRC Rate 1

SMV Rate 1 Type 0 to EVRC Rate  $\frac{1}{2}$

SMV Rate 1 Type 1 to EVRC Rate  $\frac{1}{2}$

SMV Rate  $\frac{1}{2}$  Type 0 to EVRC Rate  $\frac{1}{2}$

SMV Rate  $\frac{1}{2}$  Type 1 to EVRC Rate  $\frac{1}{2}$

....

and so on.

[0059] Each mapping module comprises a speech spectral parameter mapping unit 910, an excitation mapping unit 920, and a mapping strategy decision unit 930. The speech spectral parameter mapping unit 910 maps the EVRC line spectral frequencies directly to SMV line spectral frequencies. This occurs for all source EVRC bit rates. The parameters passed to the excitation mapping unit depend on the source EVRC bit rate. For EVRC Rates 1 and  $\frac{1}{2}$ , the input CELP excitation parameters are the pitch lag, delta pitch lag (Rate 1 only), adaptive codebook gain, fixed codevectors, and fixed codebook gain. For EVRC Rate  $\frac{1}{8}$ , typically inactive frames, the input excitation parameter is the frame energy. The excitation parameters are mapped to SMV excitation parameters, depending on the selected mapping module and

mapping strategy. The mapping strategy decision module 930 controls the mapping strategy to be used. In this example, the mapping strategy for active speech is to perform analysis in the excitation domain.

[0063] A second-stage switching module 1762 links the interpolation and mapping module to the SMV bitstream packing module 1750. The bitstream is packed according to the SMV frame type and bit rate 1716. One SMV output frame is produced for each EVRC input frame.